ISCAS 中国科学院软件研究所学术年会'2023 暨计算机科学国家重点实验室开放周
学术论文

基于对抗域适应的跨域需求链接
Cross-Domain Requirements Linking via Adversarial-based Domain Adaptation

常志远，李明阳*，王青*，李守斌，王俊杰
智能博弈重点实验室

45th International Conference on Software Engineering (ICSE 2023)

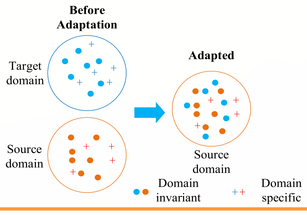
* 通讯地址: {mingyang2017,wq}@iscas.ac.cn

Introduction

- **Requirements linking:** An ability to describe and follow the life of a requirement, in both a forward and backward direction, e.g., standards to requirements and requirements to test cases.
- **Cold start projects:** 1) Some small-scale or newly started projects generally contain few requirements and linked artifacts as data resources; 2) For projects with sufficient data resources, the requirements links are typically missing or unreliable.
- **Requirements Linking Problems for Cold Start Projects**: 1) Require labeled requirements links as a prerequisite; 2) Training with sufficient labeled samples in other datasets, the distribution of features between different datasets is uneven.
- **Domain Adaptation for Cross-domain Requirements linking:** 1)Map data from different domains into the same feature space; 2)Train a linking model on a dataset with sufficient data resources (***Source Domain***), and then apply the linking model to cold-start projects (***Target Domain***) without any linking information.

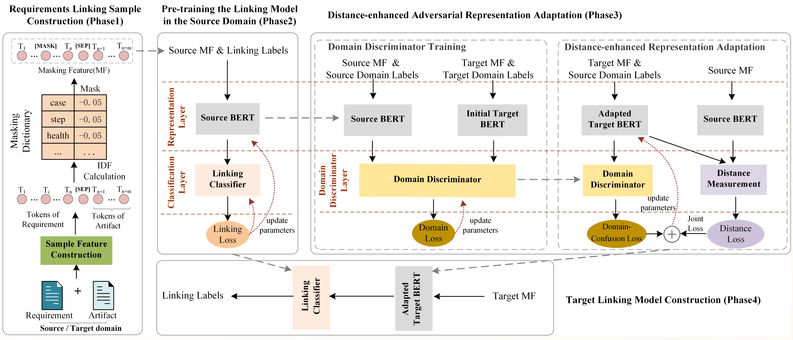
Challenge

- **Challenge-1**: Designing strategies to encourage models to prioritize ***domain-invariant features*** in cross-domain requirements linking
- **Challenge-2**: Comprehensively utilize ***adversarial-based approaches*** and ***distance measurement*** to boost cross-domain requirements linking performance.



Approach

Approach name: **RADIATION**



- **Phase 1:** Conduct standard data pre-processing, mask domain-specific features, and constructs training samples for source and target samples **(Addressing Challenge-1)**
- **Phase 2:** Pre-train a linking model for requirements linking with the labeled source samples
- **Phase 3:** Adapt the representations of the target domain to the representation space of source domain by a distance-enhanced adversarial adaptation **(Addressing Challenge-2)**
- **Phase 4:**Construct requirements linking model for the target domain

Experiment

- Dataset

Dataset	Domain	Description	True Links
HIPPA	Medical Regulations	10 regulatory codes 1889 requirements	242
CM1-NASA (CM1)	Aerospace	22 high-level requirements 53 low-level requirements	45
EasyClinic (Easy)	Management Software	30 user cases 63 test cases	63
EBT	Software System Artifacts	40 requirements 25 test cases	51
Infusion Pump (Infusion)	Medical Equipment	21 components 126 requirements	131

- Baseline

  - T-BERT: a state-of-the-art approach for the linking task by supervised learning.
  - S2Trace: a state-of-the-art unsupervised approach for requirements linking.
  - DistanceNet: a state-of-the-art approach for domain adaptation.

- Result

Metric	Model	T: HIPPA S: CM1	T: HIPPA S: Easy	T: HIPPA S: EBT	T: HIPPA S: Infusion	T: CM1 S: HIPPA	T: CM1 S: Easy	T: CM1 S: EBT	T: CM1 S: Infusion	T: Easy S: HIPPA	T: Easy S: CM1	T: Easy S: EBT	T: Easy S: Infusion	T: EBT S: HIPPA	T: EBT S: CM1	T: EBT S: Easy	T: EBT S: Infusion	T: Infusion S: HIPPA	T: Infusion S: CM1	T: Infusion S: Easy	T: Infusion S: EBT
Precision	*RADIATION*	**78.8%**	**84.2%**	**88.4%**	65.3%	**57.4%**	**67.4%**	69.8%	61.0%	**54.6%**	58.9%	59.6%	55.2%	57.1%	67.4%	**69.8%**	**61.0%**	57.4%	79.6%	75.0%	60.3%
	T-BERT (Hard-Transfer)	0.0%	33.9%	79.4%	**83.0%**	52.2%	29.3%	**77.7%**	**89.3%**	45.4%	**62.0%**	**99.5%**	**99.7%**	**62.8%**	**88.8%**	27.6%	49.5%	63.7%	**82.3%**	26.5%	70.2%
	S2Trace			44.4%				31.1%				47.2%				41.2%			33.6%		
	DistanceNet	70.2%	63.2%	54.5%	53.9%	50.0%	47.5%	45.9%	49.3%	53.0%	53.5%	53.5%	53.0%	51.8%	57.9%	58.1%	47.3%	**67.9%**	80.8%	**75.5%**	**80.5%**
	T-BERT (Within-Target)			67.7% (**60%**)				60.0% (**90%**)				66.7% (**60%**)				75.0% (**100%**)				58.3% (**100%**)	
Recall	*RADIATION*	**72.4%**	**77.6%**	59.4%	94.3%	94.6%	**89.2%**	**81.1%**	97.3%	**99.8%**	**99.9%**	**99.9%**	**99.8%**	**99.9%**	**92.9%**	**92.9%**	97.6%	**86.6%**	**80.4%**	**85.7%**	**83.0%**
	T-BERT (Hard-Transfer)	0.0%	15.2%	11.1%	18.1%	51.1%	26.6%	46.6%	55.5%	7.9%	57.1%	26.9%	36.5%	86.2%	15.7%	25.5%	**99.7%**	33.6%	10.7%	12.9%	19.8%
	S2Trace			13.1%				48.3%				73.7%				21.4%			16.8%		
	DistanceNet	44.3%	77.0%	**76.0%**	**99.5%**	**97.3%**	78.4%	75.6%	**99.7%**	99.7%	99.8%	99.7%	99.7%	99.8%	95.2%	76.1%	83.3%	79.4%	49.1%	63.4%	51.8%
	T-BERT (Within-Target)			84.0% (**60%**)				99.8% (**90%**)				80.0% (**60%**)				75.0% (**100%**)				99.8% (**100%**)	
F1	*RADIATION*	**77.9%**	**78.2%**	**71.0%**	**77.2%**	**71.4%**	**76.7%**	**75.0%**	**75.8%**	**70.6%**	**74.1%**	**74.6%**	**71.1%**	**72.7%**	**85.7%**	**82.9%**	**74.5%**	**76.7%**	**79.3%**	**79.0%**	**78.5%**
	T-BERT (Hard-Transfer)	0.0%	21.0%	19.5%	29.7%	51.7%	27.9%	58.3%	68.5%	13.5%	59.5%	42.5%	53.5%	**72.7%**	26.6%	26.5%	66.2%	44.0%	18.9%	17.4%	30.9%
	S2Trace			20.2%				37.8%				57.5%				28.1%			22.4%		
	DistanceNet	54.3%	69.5%	63.5%	69.9%	66.0%	59.1%	57.1%	66.0%	69.3%	69.7%	69.7%	69.3%	69.3%	72.0%	65.9%	60.3%	73.2%	61.1%	68.9%	63.0%
	T-BERT (Within-Target)			75.0% (**60%**)				75.0% (**90%**)				72.7% (**60%**)				75.0% (**100%**)				73.9% (**100%**)	