Putting Them under Microscope: A Fine-grained Approach for Detecting Redundant Test Cases in Natural Language

常志远，李明阳，王俊杰，王青，李守斌

The 21st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2022)
联系方式：常志远 zhiyuan2019@iscas.ac.cn
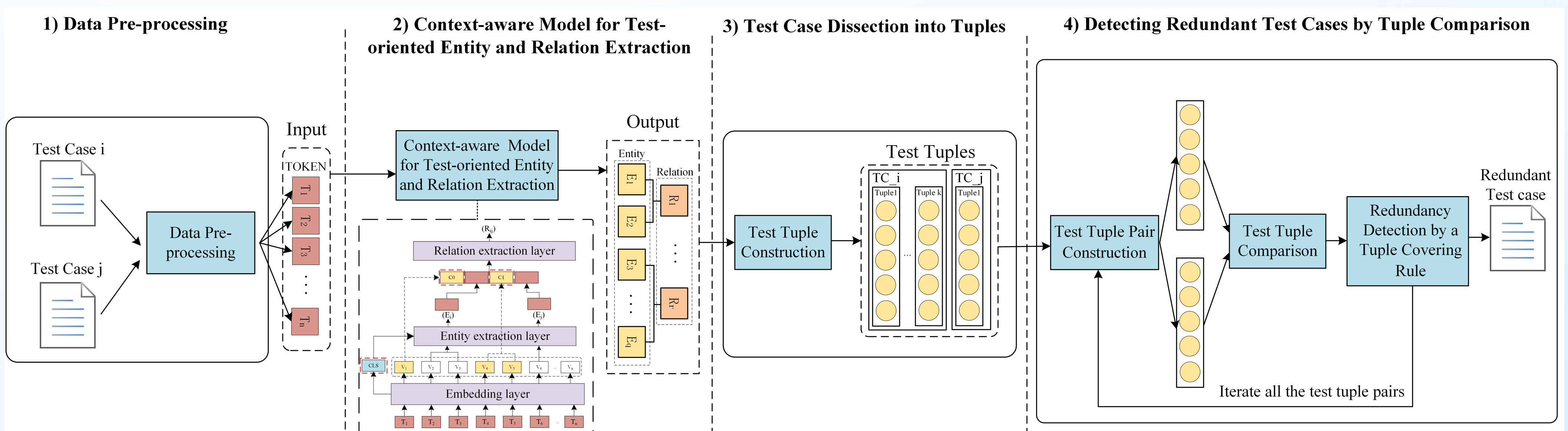李明阳 mingyang2017@iscas.ac.cn

Introduction

- As the system evolves, the redundant test cases significantly increase the cost of testing as well as maintenance efforts. Several information retrieval based approaches have been proposed to automatically detect redundancy among the NL test cases. However, these existing approaches suffer from low accuracy because they treat test cases' textual descriptions as a whole, thus can not capture its fine-grained semantic information and inherent meaning.
- There are two observations which can facilitate the similarity comparison and redundancy detection of the NL test case: 1) test case has explicit categories of test-oriented entities which can facilitate the accurate redundancy detection; 2) there might be multiple test-oriented entities that need to be carefully parsed and matched to ensure the accurate redundancy detection.
- We propose a fine-grained redundant test case detection approach Tscope, which dissects the test case into atomic test tuple(s) with the five entities restricted by their associated relations, and conducts the comparison on them. Evaluation on 3,467 test cases from ten projects shows Tscope significantly outperforming state-of-the-art approaches.

Categories

	Category	Definition	Examples
Entity	Component (Com)	the tested functional component	
	Behavior (Beh)	the behavior acting on the tested component	
	Prerequisite (Pre)	the pre-conditions before the test case is executed	
	Manner (Man)	the executing manner of the test case	
	Constraint (Con)	the constraints to be met after the test case is executed	
Relation	Act (Act)	the relation between *Component* and *Behavior*	
	Require (Req)	the relation between *Component* and *Prerequisite*	
	Use (Use)	the relation between *Component* and *Manner*	
	Satisfy (Sat)	the relation between *Component* and *Constraint*	

- Definition of entity and relation categories.
- Examples of application of entity and relation categories.

Approach

1) Data Pre-processing
2) Context-aware Model for Test-oriented Entity and Relation Extraction
3) Test Case Dissection into Tuples
4) Detecting Redundant Test Cases by Tuple Comparison

- **Data Pre-processing:** It conducts data processing and constructs samples for the extraction model.
- **Context-aware Model for Test-oriented Entity and Relation Extraction:** It designs a context-aware extraction model to extract the test-oriented entities and relations from test case descriptions.
- **Test Case Dissection into Tuples:** It dissects each test case into test tuples based on the extracted entities and relations, to represent the fine-grained test-oriented operational information.
- **Detecting Redundant Test Cases by Tuple Comparison:** It designs three comparison strategies for tuple comparison and detects redundancy by a Tuple Covering Rule.

Experiment

- Entity and Relation Extraction Performance

Metric	Model	Entity Categories					Relation Categories			
		Com	Beh	Pre	Man	Con	Act	Require	Use	Satisfy
Precision	Tscope	99.1%	99.2%	94.4%	97.3%	97.8%	90.3%	91.2%	90.1%	90.2%
	SLM	98.8%	95.2%	75.1%	91.8%	94.9%	72.8%	69.2%	85.3%	78.1%
	BLM	91.2%	88.7%	73.9%	83.1%	85.2%	84.1%	69.8%	90.1%	81.2%
Recall	Tscope	95.2%	96.9%	93.2%	95.1%	93.8%	97.9%	99.3%	97.2%	96.1%
	SLM	93.2%	92.1%	98.9%	93.2%	88.1%	94.8%	98.8%	95.1%	96.0%
	BLM	94.2%	90.9%	82.2%	84.8%	90.1%	77.2%	67.1%	65.9%	76.2%
F1	Tscope	97.1%	98.0%	93.8%	96.2%	95.7%	93.9%	95.1%	93.5%	93.1%
	SLM	95.9%	93.6%	85.3%	92.5%	91.3%	82.3%	81.3%	87.9%	86.1%
	BLM	92.6%	89.7%	77.8%	83.9%	87.5%	80.5%	68.5%	76.1%	78.6%

- Entity Category Effectiveness

Experiment Group	Precision	Recall	F1
Tscope	91.8%	74.8%	82.4%
Tscope - Com	52.2% (-39.6%)	32.7% (-42.1%)	40.2% (-42.2%)
Tscope - Beh	66.6% (-25.2%)	67.0% (-7.8%)	67.2% (-15.2%)
Tscope - Pre	86.3% (-5.5%)	67.0% (-7.8%)	75.2% (-7.2%)
Tscope - Man	84.1% (-7.7%)	65.5% (-9.3%)	73.8% (-8.6%)
Tscope - Con	83.3% (-8.5%)	66.4% (-8.4%)	73.8% (-8.6%)

- Redundancy Detection Performance