



Universiteit
Leiden

The Netherlands

Trustworthy anomaly detection for smart manufacturing

Li, Z.

Citation

Li, Z. (2025, May 1). *Trustworthy anomaly detection for smart manufacturing*. *SIKS Dissertation Series*. Retrieved from <https://hdl.handle.net/1887/4239055>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4239055>

Note: To cite this publication please use the final published version (if applicable).

Contents

Abstract	iii
1 Introduction	1
1.1 Smart Manufacturing	2
1.1.1 Industry 4.0 and Smart Manufacturing	2
1.1.2 Cyber-Physical Systems and Digital Twin	3
1.1.3 Predictive Maintenance	4
1.2 Trustworthy Anomaly Detection	8
1.2.1 Anomaly Detection	8
1.2.2 Trustworthy Artificial Intelligence	10
1.2.3 Trustworthy Anomaly Detection in the Context of Smart Manufacturing	11
1.3 Model-Centric AI and Data-Centric AI	13
1.4 Research Questions and Contributions	14
1.5 Outline of This Dissertation	19
1.6 List of Publications	20
2 Graph Neural Networks based Log Anomaly Detection and Explanation	23
2.1 Introduction	25
2.2 Related Work	27
2.2.1 Graph Representation Learning	27
2.2.2 Log Anomaly Detection and Explanation	28
2.3 Problem Statement	29
2.3.1 Graph-based Log Anomaly Detection	30
2.4 Preliminaries: Digraph Inception Convolutional Nets	31

Contents

2.5	Graph-Based Anomaly Detection for Event Logs	33
2.5.1	Graph Construction	34
2.5.2	OCDiGCN: One-Class Digraph Inception Convolutional Nets .	35
2.5.3	Anomaly Explanation	37
2.6	Experiments	38
2.6.1	Experiment Setup	39
2.6.2	Model Implementation and Configuration	40
2.6.3	Comparison to the State Of The Art	41
2.6.4	Directed vs. Undirected Graphs	42
2.6.5	Node Labels vs. Node Attributes	43
2.6.6	Robustness to Contamination	43
2.6.7	Ability to Detect Structural Anomalies and Recognize Unseen Normal Instances	43
2.6.8	Anomaly Explanation	45
2.6.9	Sensitivity Analysis	45
2.6.10	Runtime Analysis	47
2.7	An Ongoing Case Study to Lithography Systems	48
2.7.1	Background	48
2.7.2	Deployment of <i>Logs2Graphs</i>	49
2.8	Threats to Validity	49
2.9	Conclusions	50
3	Feature Selection for Fault Detection and Prediction based on Event Log Analysis	51
3.1	Introduction	53
3.2	Related Work	55
3.3	Method	56
3.3.1	Terminology and Problem Formulation	56
3.3.2	Log Event Vectorization (Module 1)	57
3.3.3	Selection of Relevant Features (Module 2)	58
3.3.4	Removal of Redundant Features (Module 3)	64
3.4	Experiments and Results	64
3.4.1	Data Description	64
3.4.2	Baseline, Hyperparameter settings and Performance metrics . .	66
3.4.3	Results and Analysis	66
3.5	Conclusion and Future Work	66

4	Explainable Contextual Anomaly Detection using Quantile Regression Forests	69
4.1	Introduction	71
4.2	Related Work	75
4.2.1	Contextual Anomaly Detection and Explanation	75
4.2.2	Traditional Anomaly Detection	77
4.3	Contextual Anomaly Detection and Explanation	78
4.3.1	Problem Statement	80
4.3.2	Overall Approach	81
4.4	Preliminaries	83
4.4.1	From Conditional Mean to Conditional Quantiles	83
4.4.2	Quantile Regression Forests	84
4.5	Quantile-based Contextual Anomaly Detection and Explanation	85
4.5.1	Detecting Anomalies with Quantile Regression Forests	87
4.5.2	Explaining Anomalies with Anomaly Beanplots	90
4.6	Experiments	91
4.6.1	Data	91
4.6.2	Baseline Algorithms and Implementations	96
4.6.3	Evaluation Criteria	99
4.6.4	Anomaly Detection Performance	99
4.6.5	Runtime Analysis	104
4.6.6	Using QCAD to Find Promising Football Players	105
4.7	Conclusions	108
4.7.1	Scaling Conditional Quantile Interval Length	115
4.7.2	Clipping Conditional Quantile Interval Length	116
5	Explainable Graph Neural Networks Under Fire	119
5.1	Introduction	121
5.2	Related Work	123
5.2.1	GNN Explanations	123
5.2.2	Attacks and Defenses of Traditional XAI Methods	123
5.2.3	Robustness/Stability of GNN Explanations	123
5.2.4	Adversarial Robustness of GNNs	125
5.3	Preliminaries	125
5.3.1	Node Classification with GNNs	125
5.3.2	GNN Explainer	126

Contents

5.4	GXAttack: <u>GNN Explanation Adversarial Attacks</u>	127
5.4.1	Attack Objectives	128
5.4.2	Relaxations and Loss Definitions	129
5.4.3	Optimization	130
5.4.4	Complexity Analysis	131
5.5	Experimental Setup	132
5.5.1	Datasets	132
5.5.2	Metrics and Baselines	133
5.5.3	Attack Transferability Settings	135
5.5.4	Training Protocols and Compute Resources	135
5.6	Experimental Results and Analysis	136
5.6.1	Experiments to Check Presumptions	136
5.6.2	Experiments to Answer RQ1 and RQ2: Attack Effectiveness . .	137
5.6.3	Experiments to Answer RQ3: Attack Transferability	138
5.6.4	Sensitivity Analysis	139
5.6.5	Property Analysis	141
5.7	Conclusions	141
6	Cross-Domain Graph Level Anomaly Detection	145
6.1	Introduction	147
6.2	Related Work	149
6.3	Problem Statement	150
6.4	Proposed Method: ARMET	151
6.5	Module 1: Graph Feature Extraction	153
6.6	Module 2: Cross-Domain Graph Level Anomaly Detection	154
6.7	Experiment Setup	157
6.7.1	Benchmark Datasets	158
6.7.2	Baselines	159
6.7.3	Evaluation	159
6.7.4	Implementation and Model Configuration	160
6.7.5	Training Hardware and Reproducibility	161
6.8	Experiment Results and Analysis	161
6.8.1	Detection Accuracy (RQ1)	161
6.8.2	Ablation Study (RQ2)	164
6.8.3	Sensitivity Analysis (RQ3)	166
6.8.4	Visualization with T-SNE	167

6.9	Discussion on Transferability	168
6.10	Conclusions	168
7	Towards Automated Self-Supervised Learning for Truly Unsuper- vised Graph Anomaly Detection	171
7.1	Introduction	173
7.2	Related Work	176
7.2.1	Anomaly Detection on Attributed Graphs	177
7.2.2	Self-Supervised Learning for Graph Anomaly Detection	177
7.2.3	Automated Self-Supervised Learning	177
7.2.4	Automated Anomaly Detection	178
7.3	Problem Statement	179
7.4	SSL for Unsupervised GAD	180
7.4.1	Existing SSL for “Unsupervised” GAD	180
7.4.2	Pitfalls in Existing Methods	181
7.4.3	Sensitivity Analysis	183
7.5	AutoGAD: Using Internal Evaluation to Automate SSL for GAD	185
7.5.1	Internal Evaluation Strategy	185
7.5.2	Discretization and Grid Search	187
7.6	Experiments	188
7.6.1	Datasets	189
7.6.2	Baselines	189
7.6.3	Evaluation Metrics	190
7.6.4	Software and Hardware	191
7.6.5	Results and Analysis	191
7.7	Alternative Strategies and Discussion	195
7.7.1	Stand-alone Internal Evaluation Strategies	196
7.7.2	Consensus-based Internal Evaluation Strategies	196
7.7.3	Discussion and Future Work	197
7.8	Conclusions	198
8	Conclusions and Future Directions	219
8.1	Conclusions	220
8.1.1	Develop and/or Improve Anomaly Detection for Smart Manu- facturing from A Data-Centric AI Perspective	220
8.1.2	Develop and/or Improve Anomaly Detection for Smart Manu- facturing from A Model-Centric AI Perspective	222

Contents

8.2	Limitations and Future Directions	225
8.2.1	Limitations of Proposed Methods	225
8.2.2	Other Future Directions	229
	Acknowledgements	271
	Summary	273
	Samenvatting	277
	SIKS Dissertation Series	281
	Curriculum Vitae	295