

Outlier Detection of Generalized Deduplication Compressed IoT Data

Something

Morten Lyng Rosenquist
Faculty of Technical Sciences
Aarhus University
Aarhus, Denmark
201706031

April 29, 2022

Abstract—

- Summary of whole report
- Write last

Index Terms—

- Buzzwords regarding field of work
- 3-8 words

I. INTRODUCTION

- Introduce the subject
- What does the paper research
- What, how and why?
- Relation to IoT
- Start broad and narrow it down to what will be looked at in paper
- Write later
- Faster computation, maintaining precision
- Operate on smaller data
- Only transmit bases and counts
- Motivate that stuff can be done on edge

II. RELATED WORK

- Should this be previous/related work?
- Does this section make sense
- In Direct Analysis of GD... there is a background section describing GD and clustering etc. Should we take that approach?
- What to include in this? GD, State of The art Anomaly Detection, Isolation Forest, etc.?
- Background makes sense for GD and Isolation Forest
- Related work state of the art
- Maybe related work part of introduction
- Include both related and background

III. METHODS

- Here we describe how the work was done.
- Mathematical description of new model. Maybe pseudocode of algorithm?

- Discussion of new model? Performance, complexity, etc?
- We describe the extended isolation forest here. In Direct Analysis of GD .. the used method has its own section.
- Concept, and new stuffs
- Running isolation as is directly on bases - phase 1
- Running extended isolation on the bases - phase 2
- Can include or exclude counts
- tradeoffs: counts, precision.

IV. EXPERIMENTS & RESULTS

- Should this be split in 2 as usual? In direct analytics of GD... there is a "performance of..." section. Is that better?
- Description of experiments
- Describe dataset
- Suboptimal extension of sklearn
- Performance: time, f1, recall, precision, accuracy.
- Comparisons (quantitative and qualitative)
- Illustration of performance (Precision-Recall, ROC, confusion matrix)

V. CONCLUSION

- Summary of the work
- Key Observations
- Limitation and future work ()