SCALING GEOMETRIC MONITORING OVER DISTRIBUTED STREAMS

by

Alexandros D. Keros

A thesis submitted in partial fulfillment
of the requirements for the degree

of

UNDERGRADUATE

in

Electronic and Computer Engineering

Approved:

Dr. Vasilis Samoladas Major Professor	first reader Committee Member
second reader Committee Member	dean

Technical University of Crete
Chania, Crete, Greece

2015

Scaling Geometric Monitoring over Distributed Streams

by

Alexandros D. Keros, Undergraduate

Technical University of Crete, 2015

Abstract

BLAH BLAH

Thesis Supervisor: Dr. Vasilis Samoladas
Department: Electronic and Computer Engineering

(24 pages)

Public Abstract

BLAH BLAH

Acknowledgments

my mum

Contents

List of Tables

Table Page

List of Figures

Figure Page

Part I

INTRODUCTION AND PRELIMINARIES

Chapter 1

Introduction

1.1 Overview

1.2 Motivation

1.3 Contributions

1.4 Thesis Outline

Chapter 2

Theoretical Background

The present chapter contains the necessary background knowledge used throughout the length of this thesis. Section 2.1 describes the *"Geometric Approach to Monitoring Threshold Functions over Distributed Data Streams"* in detail, as formulated by I.Sharfman, A.Shuster, D.Keren [**?**]. Section 2.2 presents *multi-objective optimization* and dives into the algorithms used in our implementation. Section 2.3 discusses *graph maximum weight matching* used for node pairing and, finally, in Section 2.4 we explain the *Savitzky-Golay filtering* used for velocity approximation.

2.1 Geometric Monitoring of Distributed Streams

The Continuous Distributed Monitoring Model, a.k.a. Data Stream System

idea: having a real-time overview over the system

application examples: ISP network traffic, distributed sensors etc

complexity: monitoring value or threshold monitoring over the whole set of observations, in real time

goal: minimize communication while retaining the highest accuracy possible

possible solutions:

1.centralize

- suffers from network overload, storage overload

2.poll

- not real time, update frequency-accuracy trade-off

3.GM monitoring

-apply convex opt theory in order to reduce communication while retaining accuracy bounds

details of geometric monitoring model

2.1.1 System Architecture

fully distributed node topology

.no coordinator-center node

.communication between nodes

image

coordinator based node topology

.coordinator-center node

.nodes communicate only with coordinator

image

2.1.2 Computational Model

stream and node notation

weights

statistics vectors

global statistics vector

monitored function

threshold

estimate vector

drift vector

general operation of distributed algorithm

drift vector definition

general operation of coordinator based algorithm

balancing process

slack vector

drift vector definition

2.1.3 Geometric Interpretation

node local constraints make sure global violation is accurately monitored

how?

convexity property of drift vectors

theorem of bounded convex hull by local constraints (balls)

monochromaticity of balls

balls monochromatic means threshold upheld

2.1.4 Protocol

decentralized algorithm (in short, for completeness)

centralized algorithm (in detail)

we will focus on that

2.2 Multiobjective Optimization

what is mop

use examples

kinds:

a.numerical

b.evolutionary

2.2.1 Sohr's algorithm a.k.a. ralg

algorithm description

2.3 Maximum Weight Matching in Graphs

general graph theory (introductory)

what is max weight matching

algorithm description

2.4 Savitzky-Golay Filtering

filtering generals

examples of uses of filters

filters:

Kalman

+,- Moving Average

+,- Savitzky-Golay a.k.a. ??? +,-

algorithm description

Chapter 3

Related Work

Part II

PROBLEM DEFINITION AND IMPLEMENTATION

Chapter 4

Problem Statement

Chapter 5

Implementation

5.1 Geometric Monitoring

5.2 Heuristic Balancing

5.3 Distance Based Node Matching

5.4 Implementation Challenges

Part III

RESULTS AND CONCLUSIONS

Chapter 6

Experimental Results

6.1 Experimental Setting

6.2 Heuristic Balancing

6.3 Distance Based Node Matching

6.4 Overall Results

Chapter 7

Conclusions and Future Work

7.1 Conclusions

7.2 Future Work

References

Appendix

Chapter A

Geometric Monitoring Python Implementation

A.1 Python

A.2 Numpy and Scipy

A.3 Openopt

A.4 NetworkX

A.5 Putting It All Together