PRESENTATION OUTLINE: A parallel implementation of outlier detection over data streams

Kangqing YU
School of Computer Science
Carleton University
Ottawa, Canada K1S 5B6
kangqingyu@cmail.carleton.ca

November 20, 2017

1   Outlier detection task

- Outlier detection definition

- Typical methods & algorithm

- Main application

2   Data streams

- Characteristics of data streams

- Requirements of real-time data streams processing

- Outlier detection over data streams

- Main challenges

3   Introduction to parallel computing

- GPU and SIMD

- NVIDIA & CUDA framework

- Others

4   Goal of this project

- Intersection of outlier detection, data streams and parallel computations

- Brief description of proposed approach

- Novelty

5 Local Outlier Factor(LOF) algorithm

- General idea

- Steps and mathematical formulas

- Extremely computational expensive

6 Statistical binned summary

- What is contained in the binned summary

- How to calculate the bin index for each data point

- How to maintain/update the binned summary

- Curse of dimension problem

7 Cumulative LOF algorithm(modified version)

- General idea

- Role of the binned summary

- Steps and mathematical formulas

8 Implementation of cumulative LOF on CUDA

- Coalesced memory access & shared memory

- Detailed steps in host and kernel functions

- Threads managements in each kernels

9 Binned summary maintenance using MapReduce

- Why MapReduce?

- Map stage

- Reduce stage

10 Q & A

- Questions?