Accelerating Anomaly Detection Algorithms on FPGA-Based High-Speed NICs

Hiroki Matsutani
Dept. of ICS, Keio University
http://www.arc.ics.keio.ac.jp/~matutani
Accelerator design for big data

Input stream data

Message queuing

Stream processing

Inference

Database layer (Polyglot persistence)

Realtime View

Batch View

Data exchange (serialization)

Batch processing

Batch Learning

Big data (Surveillance, Network service, SNS, UAV, IoT)

DB queries

✓ Customer analysis
✓ Topic prediction
✓ Blockchain records
✓ Geolocation query
✓ …
Today’s talk: Online learning

Input stream data

- Big data (Surveillance, Network service, SNS, UAV, IoT)

Message queuing

- Stream processing
  - Inference, Online learning
  - Batch Learning

Database layer (Polyglot persistence)

- Realtime View
  - DB queries
  - Customer analysis
  - Topic prediction
  - Blockchain records
  - Geolocation query
  - …

- Batch View
  - Batch processing

Data exchange (serialization)

- Data exchange (serialization)

I/O intensive

- Message queuing middleware (Apache Kafka)
- Stream processing (Apache Spark Streaming)
- Inference (DNN, CNN, RNN)
- Online learning (SGD, ChangeFinder, OS-ELM)

Compute intensive

- Document DB (MongoDB)
- Graph DB (Neo4j), graph processing
- Batch learning (Apache Spark)
- Batch processing (Apache Spark)
- Multi-Processing System on Chip (MPSoc)

Tight integration of I/O and compute \(\rightarrow\) FPGA

Massive parallelism \(\rightarrow\) Networked GPU cluster
Offline vs. Online learning

- Examples: DNN, CNN, …
- Learning cost is high
- Predictor updated infrequently

Offline learning

- All training data
- New training data
- GPU-based batch processing
- OK or NG?
- Test data

Online learning

- Learning cost is low
- Predictor updated frequently
- Not very versatile

Sequential learning + Inference

- OK or NG?
- Predictor
- Test data ≒ Training data

FPGA-based stream processing

FPGA NIC/Switch

10GbE x4
Online learning approaches

ChangeFinder:
Outlier and change point detections on time-series data

Next value $X_t$ is predicted based on recent $p$ values

AR-model based

Neural network

Online sequential learning for SLFN (input, hidden, and output layers)

\[ X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots \]
ChangeFinder on 10GbE FPGA

• ChangeFinder algorithm [J. Takeuchi, IEEE TKDE'06]

**Step 1 (Outlier score):**
Receive input data $X_t$ at time $t$
Calculate outlier score of $X_t$ based on past data
Influence of past data controlled by discount rate $r$

**Step 2 (Smoothing):**
Calculate moving average $Y_t$ of the outlier score
Smoothing is controlled by window size $S$

**Step 3 (Change-point score):**
Step 1 is performed for $Y_t$
The result is change-point score

Input data $X_t$

Change-point score

Elapsed time
ChangeFinder on 10GbE FPGA

- ChangeFinder algorithm [J. Takeuchi, IEEE TKDE’06]

**Step 1 (Outlier score):**
Receive input data $X_t$ at time $t$
Calculate outlier score of $X_t$ based on past data
Influence of past data controlled by discount rate $r$

**Step 2 (Smoothing):**
Calculate moving average $Y_t$ of the outlier score
Smoothing is controlled by window size $S$

**Step 3 (Change-point score):**
Step 1 is performed for $Y_t$
The result is change-point score
ChangeFinder on 10GbE FPGA

- 10GbE NIC datapath by Verilog HDL
- Application logic in wrapper in HLS

[T.Iwata, HeteroPar’18]
ChangeFinder on 10GbE FPGA

• Throughput: 83.4% of 10GbE line rate

Youtube Video: https://www.youtube.com/watch?v=wgTcBfkE5hY

The ChangePoint score (left) becomes large when the time series data (right) is changed.
Online learning approaches

**ChangeFinder:**
Outlier and change point detections on time-series data

**AR-model based**
Next value $X_t$ is predicted based on recent $p$ values

**OS-ELM:**
Single hidden layer neural network (SLFN)

**Online sequential learning for SLFN (input, hidden, and output layers)**

$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots$

**Weight vector**
$\beta_{00} \sim \beta_{(N-1)(m-1)}$

**ChangeFinder:**
Outlier and change point detections on time-series data

**AR-model based**
Next value $X_t$ is predicted based on recent $p$ values

**OS-ELM:**
Single hidden layer neural network (SLFN)

**Online sequential learning for SLFN (input, hidden, and output layers)**

$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots$

**Weight vector**
$\beta_{00} \sim \beta_{(N-1)(m-1)}$
Online learning + unsupervised

Offline learning

Inference only

Test data ➔ Pre-trained predictor

OK or NG?

Online learning

Sequential learning + Inference

Learning predictor

Test data ≈ Training data

Unsupervised anomaly detection (No training data needed)

[M.Tsukada, HeteroPar’18]
*Collaboration with Prof. M.Kondo (UTokyo)

Normal values (incl. noise) are learned after the deployment ➔ Anomaly detection adapted to a given environment
Online learning + unsupervised

• Learn vibration pattern of fan + noise

Youtube Video:
https://www.youtube.com/watch?v=tCw7p7bjwTs

Vibration pattern is changed because of the air spray.
→ Detected as "Anomaly"
Summary: Online learning FPGA

Input stream data

Message queuing

Stream processing

Inference, Online learning

Database layer
(Polyglot persistence)

Realtime View

DB queries

- Customer analysis
- Topic prediction
- Blockchain records
- Geolocation query
- ...

Batch View

Big data (Surveillance, Network service, SNS, UAV, IoT)

Data exchange (serialization)

Batch Learning

Batch processing

Input stream data

Message queuing

Stream processing

Inference, Online learning

Database layer
(Polyglot persistence)

Realtime View

Big data (Surveillance, Network service, SNS, UAV, IoT)

Data exchange (serialization)

Batch Learning

Batch processing

I/O intensive

Message queuing middleware (Apache Kafka)

Stream processing (Apache Spark Streaming)

Inference (DNN, CNN, RNN)

Batch processing (Apache Spark)

Compute intensive

Serialization (Apache Thrift)

KVS / Column DB (Redis, HBase)

Online learning (SGD, ChangeFinder, OS-ELM)

Graph DB (Neo4j), graph processing

Tight integration of I/O and compute \( \rightarrow \) FPGA

Massive parallelism \( \rightarrow \) Networked GPU cluster

GPUs

Switch

Host

Today’s talk

MPSoc16

MPSoc17

MPSoc16

Four 10GbE

FPGA
• Outlier detection on 10GbE FPGA NIC

• Change-point detection on FPGA NIC
References (2/2)

• Online sequential unsupervised anomaly detector on FPGA
Thank you for listening

Acknowledgement:
This work is supported by JST CREST JPMJCR1785