

A FOG COMPUTING PROTOTYPE

Course Project for Big Data Analytics — Winter 2019

Ali Alizadeh Mansouri Marco Sassano

April 15, 2019

Concordia University

INTRODUCTION

WHAT IS FOG COMPUTING?

Internet of Things (IoT)

- 50 billion devices (sensors, smart phones, smart cities, healthcare, smart vehicles, ...)[5]
- 100s of terabytes towards petabytes per day
- limited computing resources

WHAT IS FOG COMPUTING?

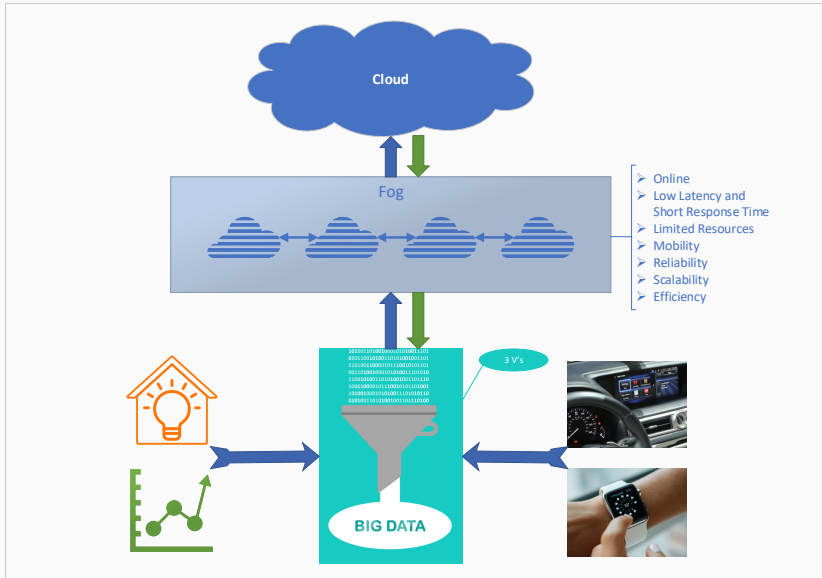
Internet of Things (IoT)

- 50 billion devices (sensors, smart phones, smart cities, healthcare, smart vehicles, ...)[5]
- 100s of terabytes towards petabytes per day
- limited computing resources

Cloud computing

- Flexible economy
- Scalability
- Adaptation to varieties of computational demands
- Suffers from high latency

A FOG COMPUTING ENVIRONMENT



PROBLEM

Early detection of epilepsy seizures using EEG timeseries data

Inspired by Diab Abdulgalil et. al. [4]

- Stream processing on the edge
 - Lower accuracy, but very high response time
- Clustering in the cloud
 - Longer latency and response time, but more accurate

DATASET

- Original dataset[2]: 500 individuals, each with 4097 data points for 23.5s
- The UCI Machine Learning Repository version [1]:
 $23 \times 500 = 11500$ records, each record contains 178 data points (columns) for 1 second
- Restructured to (time, value) tuples to be used as a stream
 $\Rightarrow 11500 \times 178 = 2\,047\,000$ tuples (windows of 178)
- 5 classes \rightarrow 2 classes (binary)

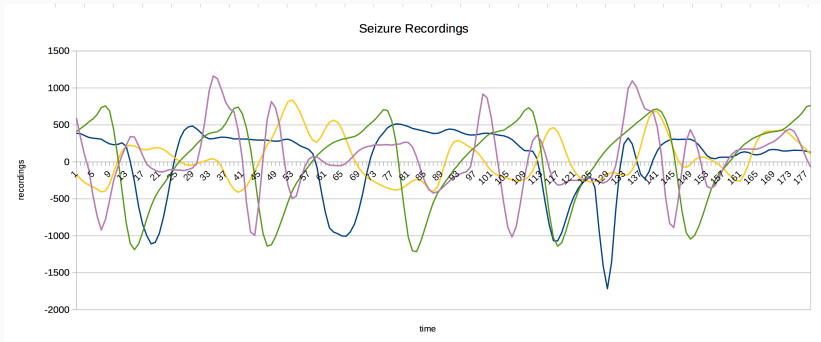
	Total	Positive	Negative
Training	7 000	1 392	5 608
Test	4 500	908	3 592
	11 500	2 300	9 200

METHODOLOGY

STREAM FILTERING

Goal Filter the stream for out-of-range tuples/anomalies

- light-weight on resources
- fast



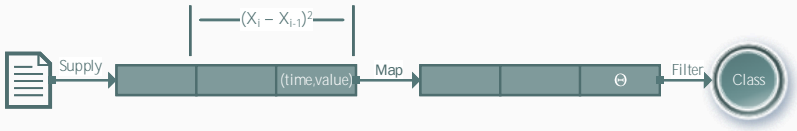
Different statistical measures: variance, skewness, kurtosis [3]

Selected measure

$$\Theta = \sqrt{\sum_{i=2}^n (X_i - X_{i-1})^2}$$

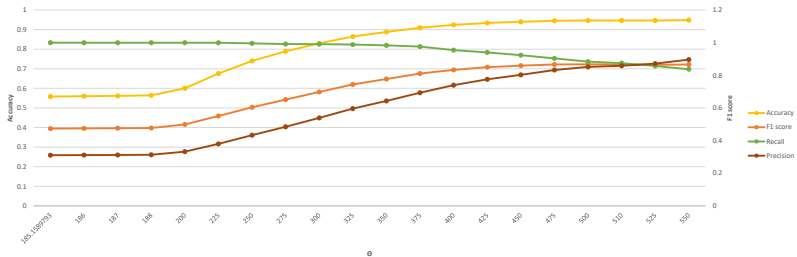
Apache Edgent: Programming model and runtime (Java)

- **Supply** reads the dataset and supplies it as a stream
- **Map** maps the tuples to the selected measure
- **Analysis** classifies the out-of-range Θ values as seizures

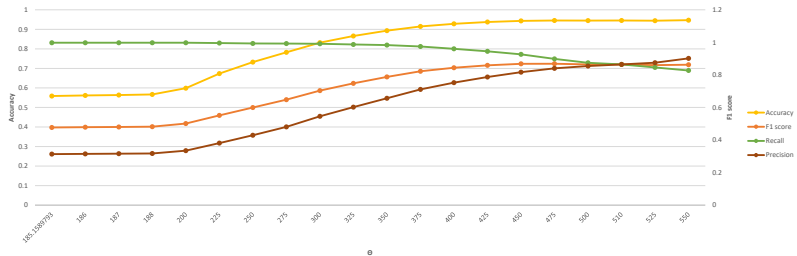


RESULTS

TRAINING DATA



TEST DATA



Inspirations

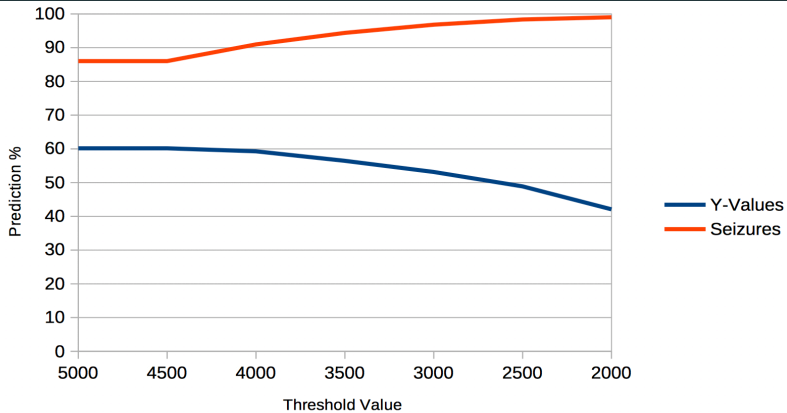
- The EGADS anomaly detection library by Yahoo[6] uses clustering as basis for most of its
- Carney et. al.[3]

Apache Spark (pyspark/Python)

Selected distance measure

$$s = \sum_{i=2}^n |X_i - X_{i-1}|$$

RESULTS



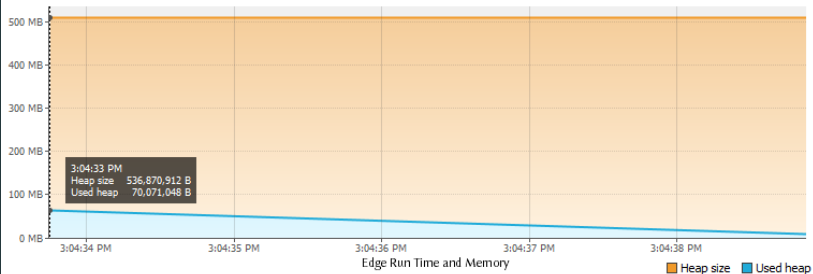
CONCLUSION

PERFORMANCE COMPARISON

Size: 536,870,912 B

Used: 11,878,680 B

Max: 8,575,254,528 B



- Experiment with more than one sensor
- Introduce a publish-subscribe platform such as Apache Kafka, IBM Watson IoT Hub, etc. (e.g. using MQTT as the protocol)

Thank you!



Epileptic seizure recognition data set.

**[https://archive.ics.uci.edu/ml/datasets/
Epileptic+Seizure+Recognition](https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition).**

Accessed: 2019-04-15.



R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David,
and C. E. Elger.

**Indications of nonlinear deterministic and
finite-dimensional structures in time series of brain
electrical activity: Dependence on recording region and
brain state.**

Phys. Rev. E, 64:061907, Nov 2001.



P. R. Carney, S. Myers, and J. D. Geyer.

Seizure prediction: methods.

Epilepsy & behavior : E&B, 22 Suppl 1(Suppl 1):S94–S101, dec 2011.



H. Diab Abdulgalil.

A Multi-Tier Distributed fog-based Architecture for Early Prediction of Epileptic Seizures.

PhD thesis, University of Waterloo, 2018.



D. Evans.

The Internet of Things: How the Next Evolution of the Internet is Changing Everything.

Technical report, 2011.



N. Laptev, S. Amizadeh, and I. Flint.

Generic and scalable framework for automated time-series anomaly detection.

In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1939–1947.

ACM, 2015.